Computer–Assisted Quality of Life Assessment for Clinical Trials

Leslie A. Lenert, M.D., M.S.[1] and John C. Hornberger, M.D., M.S. [2]
Divisions of Clinical Pharmacology[1] and Health Services Research [2]
Stanford University School of Medicine, Stanford, CA 94305.

A patient's quality of life is difficult to assess. Most methods designed to evaluate quality of life for clinical trials are time consuming, subject to varied interpretation among patients and assessors, and apply scales with only a limited relationship to utility theory. We have developed a HyperCard program, incorporating animated graphics, to (1) improve the speed of collection and collation of data, (2) improve understanding between patients and assessors, and (3) incorporate utility theory-based assessments. We assessed the quality of life of 25 patients with end-stage renal disease receiving hemodialysis using a traditional paper-based presentation and, eight months later, repeated the assessment with a computer-assisted presentation. Each presentation incorporated five techniques for evaluating quality of life: the Campbell Index of Well-being, the Kaplan-Bush Index of Well-being, categorical scaling, standard gamble, and time tradeoff. The computer program improved the speed of collation of data for statistical analysis, provided a consistent interface among patients for utility assessments, and showed stable reporting of patient's well-being. These findings suggest that further development of computer-assisted assessments of patients' quality of life for clinical trials is warranted.

INTRODUCTION

Many indices have been developed to assess patients' quality of life related to underlying illness or response to therapy. These methods are being applied more frequently in clinical trials, but quality of life assessments are time-consuming and subject to varied interpretation by patients and assessors. Moreover, individual assessment of quality of life using utility theory rarely has been applied in clinical trials. The principle advantage of utility theory-based assessments is the direct comparison of distinct medical impairments based on a reasonable set of assumptions about patients preferences [1]. Specifically, utility assessments allow a comparison of anticipated benefits of treatment to risk of death by the patient or costs assumed by society. These features are particularly valuable in assessing the appropriateness of therapies that present substantial risk to patients, and for use in policy decisions on the allocation of health care technologies.

Performing utility assessments to evaluate the impact of a technology has previously not been practical. Utility assessment requires a large time commitment, and assessments must be performed by trained individuals. We have developed a HyperCard program designed to collect a broad set of quality of life estimates for patients, including utility assessments. To facilitate utility assessment, we have developed cards that use animated graphics that display the tradeoffs required for utility assessment. In this paper we will describe an initial approach to automation of utility assessments, and give the results from a preliminary study applying the program in a group of patients receiving hemodialysis for end-stage renal disease.

BACKGROUND

The components of quality of life encompass objective elements, such as the level of physical limitation in performing daily active activities; and subjective measures such as the patient sense of well-being [2]. The Kaplan-Bush Index is an example of a tool which uses objective attributes to define patients' level quality of life [3]. It uses, in part, patients' principal physical limitation to obtain an estimate of overall quality of life.

Subjective assessments of quality of life focus on patients' emotional response to their illness. A well established method is the Campbell Index of Well-being, which has been applied to broad groups of patients [4, 5]. The well-being portion of the questionnaire is brief (one page) and well correlated with the results of the overall scale. An alternative method is categorical scaling, where persons rate their current quality of life on a scale from 0 (death) to 100 (normal health life) [6]. These approaches are simple to administer, but have little foundation with theories of normative decision making.

Utility based scales of quality of life are also subjective. Utility assessments determine the value of a health state in terms of what a patient is willing to trade in terms of money, length of life, or risk of death. For serious and disabling diseases such as renal failure, stroke, or coronary artery disease, utility assessments are commonly presented in terms of willingness to tradeoff length of life to gain normal health [7-11] or willingness to risk a chance of an immediate, but painless, death to gain normal health [7-9, 12, 13]. The former method is called the time tradeoff approach, and the latter method is called the standard gamble. Estimation of quality of life with time tradeoff or standard gamble is subject to bias

0195—4210/96/$5.00

depending on how the tradeoff is framed (i.e., risk of death or chance of life) [14]. The standard gamble can be influenced by risk aversion and, for people with extreme severe impairments, by the number and type of prospects considered by the respondent, so called Prospect Theory [14]. Time tradeoff assessments may be biased by excessive discounting of future life years [10].

Assessors play an important role in estimation of quality of life using utility theory. By repeated questioning and rephrasing of the tradeoffs, assessors can determine whether patients' responses are internally consistent. Assessors also function to prevent *anchoring* of patients tradeoffs on initial estimates of utilities values and to correct other biases they perceive in the patient responses.

The Kaplan-Bush Index, categorical scaling, Campbell Index, and others have been applied more widely in clinical research than utility-based methods, primarily because they are easier to use. However, nonutility-based methods do not conform to standard theories of individual behavior of decision making. Inferences from such studies using nonutility-based methods are necessarily limited, particularly when trying to draw policy conclusions. Moreover, studies mapping utility to nonutility-based methods demonstrate marked discrepancies in the patients' perceived well-being [15].

In summary, techniques for the assessment of patients' quality of life are still evolving. Further refinements in quality of life assessments in clinical trials include improving the speed of collection or collation of data, reducing assessment biases, and incorporating utility-based methods into routine testing. We designed a computer-based assessment program to address these concerns.

DESIGN CONSIDERATIONS

A program for assessing quality of life must provide tools that can be used to evaluate a broad range of patients in a variety of study settings, and not simply for a single study. In most cases, the relationship between therapy and quality of life is often specific to an area of medicine. Therefore a quality of life assessment program must be a readily adaptable tool.

Categorical scaling, Campbell Index, and Kaplan-Bush Index are traditionally implemented by questionnaire on paper media. To allow comparison of results with previous studies, it is desirable that the electronic representation be a close approximation of the paper-based method.

In contrast, the use of paper media for utility assessment has been problematic. Text descriptions of probabilities or time tradeoffs may not be readily envisioned by patients. Anchoring on the first number conceived by the individual is likely to occur [16, 17]. To aid in assessing probabilities, such as is required by the standard gamble, visual references are generally used by the assessor. The visual aid most frequently used is a probability wheel, in which a pie chart is used to display the risk of death that the patient is willing to accept. For the estimation of the probabilities of death of less than ten percent -- the range of risk that most patients are willing to assume -- human perception may limit the usefulness of the probability wheel [13].

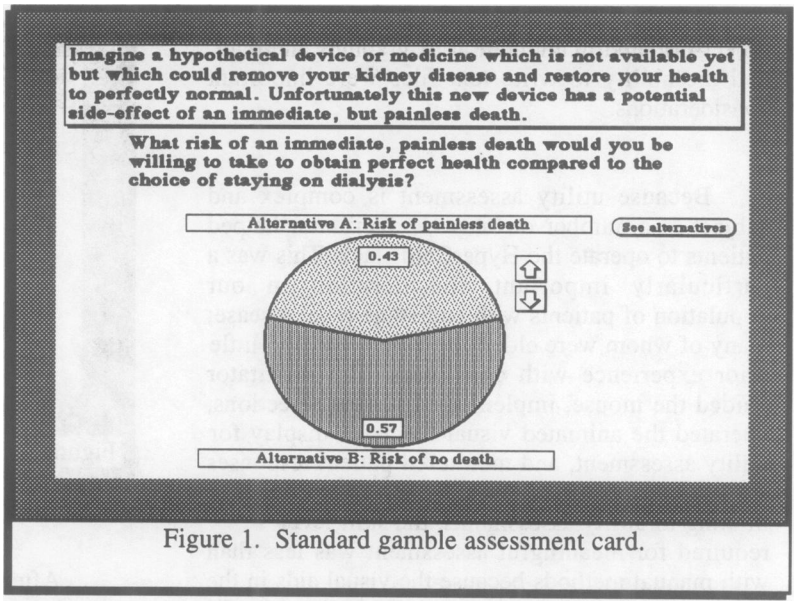

Figure 1. Standard gamble assessment card.

Similarly, performing time tradeoffs may present specific problems. First, determining the life expectancy of individual patients may be difficult because of the presence of comorbid diseases. Second, presentation of a realistic life expectancy may upset patients with short life expectancies. Third, patients are likely to discount future life years as they estimate their willingness to trade life expectancy for better health. Like standard gamble, visual references have been used to improve time tradeoff assessments [18].

IMPLEMENTATION

We chose HyperCard to implement our quality of life assessment program. HyperCard allowed us to build a template stack with a general design framework that could be rapidly modified for specific trials. Ease of program modification, access to

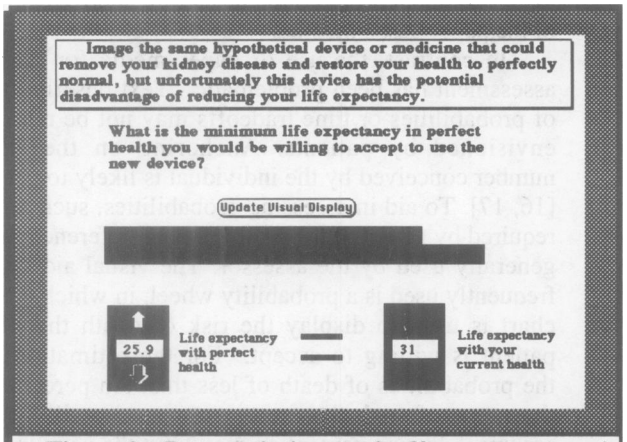

Figure 2. Bar graph time tradeoff assessment card.

object-oriented manipulation of graphics and data, and the ability to write text files were important considerations.

Because utility assessment is complex and subject to a number of biases, a facilitator helped patients to operate the HyperCard stack. This was a particularly important consideration in our population of patients with end-stage renal disease; many of whom were elderly and frail, and had little prior experience with computers. The facilitator guided the mouse, implemented patient selections, operated the animated visual feedback display for utility assessment, and monitored patient responses for internal consistency. While the facilitator needed training in utility assessments, the skill level required for meaningful assessment was less than with manual methods because the visual aids in the computer- assisted program provided a meaningful symbol of the tradeoff that patients were asked to make.

Cards for assessing the Kaplan-Bush Index, categorical scaling, and Campbell Index were designed to be exact duplicates of standard paper forms used in clinical trials, except the mouse and key board were substituted for pen and paper. Our initial design for assessment of utilities by the standard gamble method used an animated probability wheel (Figure 1). Patients were first introduced to the idea of a standard gamble in an initial card offering trades of small amounts of money. On a second card, patients were asked to assess the risk of death they would accept to return to normal health compared their current health state. Alternatives were hidden on initial presentation of the card, and then revealed by the facilitator.

The first design for estimation of utilities by the time tradeoff approach was implemented using one- dimensional comparison of length of life (Figure 2.). A subsequent design used a two-dimension comparison of length and time tradeoff, with length of life on the horizontal axis and quality of life on the vertical axis. Patients traded life expectancy for quality of life. This design is shown in Figure 3. The two-dimensional display was incorporated into the renal failure assessment program because the rectangular illustration provides a vivid demonstration that later years of life should not be discounted in the assessment; every year of life should be considered equivalently by the patient.

Results from various assessments are stored in global variables. These variables are collected into a single field in the last card of the HyperCard stack, which is displayed so that the facilitator can validate the data entered prior to saving it or printing it.

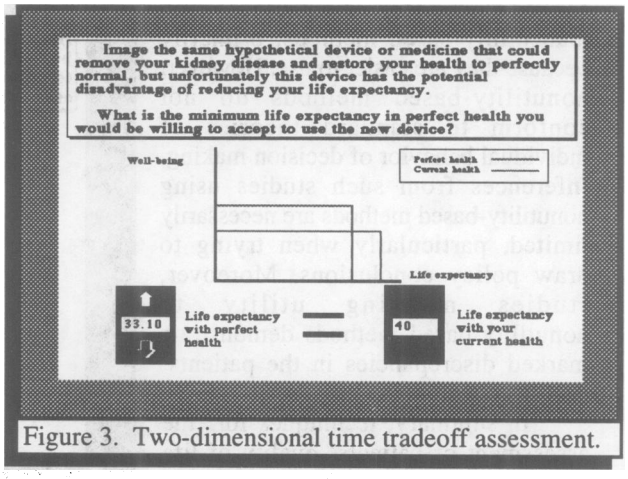

Figure 3. Two-dimensional time tradeoff assessment.

PROGRAM STATUS

After development of the template HyperCard stack, we adapted the stack to two different assessment tasks: assessment of quality of life in patients with renal failure and impact of drug therapy on quality of life of patients with mental illness. The program for quality of life assessment in mental illness is still in development. Assessing patients with a moderate degree of mental illness poses unique challenges. Tests of items discussed in the program had to be incorporated to ensure patients understood the meaning of questions. One patient with schizophrenia had a quality of life assessment using the program. After early testing, further design work was initiated.

The renal failure program has been used to assess quality of life in 43 patients with chronic renal failure on dialysis. In a pilot study, we compared the program with a paper-medium survey of quality of life. Patients from the Stanford University Hospital Dialysis Center previously completed a paper-based survey that assessed quality of life using the

	CI	KB	CS	SG	TTO
paper	64	65	68	66	71
s.d.	(+/-25)	(+/-13)	(+/-28)	(+/-34)	(+/-20)
comp	70	70	62	72	84
s.d.	(+/-23)	(+/-19)	(+/-18)	(+/-25)	(+/20)

Table 1. Mean scores for quality of life with standard deviations for paper and computer-based methods.CI = Campbell Index; Kaplan-Bush Index; CS = categorical scaling; SG = standard gamble; TTO = Time tradeoff.

Kaplan-Bush Index, categorical scaling, Campbell Index, standard gamble, and time tradeoff. Assessment methods for the standard gamble and time tradeoff in the previous study were performed as interviews. A physician with formal training with decision analytic methods performed utility assessments for the initial study. Visual reference aids for standard gamble and time tradeoff were not used at the first interview.

Eight months later, we obtained informed consent from 25 of these 43 patients who had completed the survey. The remaining patients had either died or had been discharged to a different dialysis center, so were not available for a follow-up assessment. Another physician, unfamiliar with decision analysis, was recruited to be the facilitator. She was instructed in the operation of the program, and then obtained quality of life estimates in the 25 patients.

Standard gamble estimates of utility were obtainable in 24 of 25 patients and a time tradeoff estimates in all 25 of 25 patients. The lack of a single response on standard gamble related to the patient's inability to comprehend the concept of probabilities. Patients completed all other data items. Collating data for statistical analysis from the computerized quality of life assessments was substantially easier than with paper-based assessments. There were no appreciable time savings for collecting data over paper-based methods. Mean scores of different assessments and methods are shown in table 1.

We tested the hypothesis that scores of the utility-based methods would vary between the media used, whereas scores of the nonutility-based methods would remain unchanged. The effect of assessment method on individual scales was tested using Hotelling's T^2 [18]. There was a significant difference between assessment methods only for the time tradeoff utility values ($p < 0.012$). Computer-based time tradeoff values were about 20 percent higher than had been observed in a previous study of quality of life using time tradeoff assessments in a different population of patients [19]. The reason for this discrepancy is unclear, but may be related to different patient characteristics, the visual aids used display trade-offs, or the methods used to search for subjects' indifference point in utility elicitation.

LESSONS LEARNED

Clinical trials are often conducted in multiple centers with data collected at many time intervals. As a result, it is important that the methods for obtaining data be as consistent and reliable as feasible. This computer-assisted program for assessing quality of life provides a method for data collection which (1) obtains results comparable to

The computer also served as a means of recording and storing data from the clinical trial. Data analysis was greatly improved by computer data collection. This advantage alone seemed significant enough to warrant future application of computer-based assessment of quality of life. Scores on computer-assisted assessments were comparable to paper-based methods. The role of visual feedback in quality of life assessments still needs clarification. For example, the visual display for the standard gamble may induce subjects to assume less risks than they would without visual feedback. Also, the two dimensional time tradeoff display produced estimates of patient utilities that were much higher than in our prior study or in the literature; further work is warranted to determine the most reliable visual aids for illustrating the time tradeoff and other utility decisions.

Our work with assessing quality of life in patients with mental illness, though still at a preliminary stage, taught us several valuable lessons. First, patient comprehension of text descriptions of health states cannot be taken for granted and documentation of comprehension of tradeoffs is important to ensure their validity of assessments. Second, assessing a patient's attitude about symptoms not yet experienced is difficult. Text descriptions of the symptoms of diseases or drug side

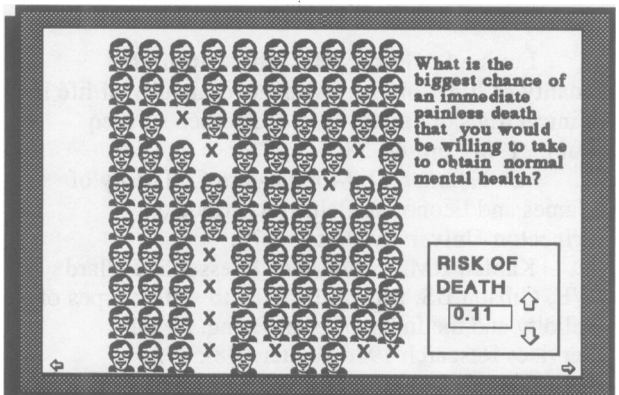

Figure 4. An alternative to probability wheel for visualizing risks of death in the Standard Gamble.

effects (such as given on the cards in figure 1-4) may be too abstract for patients to imagine for themselves without help. Text descriptions need to be supplemented or replaced with more powerful methods of communication, such as short multimedia presentations or brief video tapes defining the aspects of quality of life under study.

FUTURE PLANS

We will continue to implement test designs for computer-assisted quality of life assessment in clinical research settings, working to improve communication of difficult concepts inherent in quality of life assessment and emphasizing utility assessment because of its relationship to medical decision analysis. Our most recent design work focuses on a tool to replace the probability wheel in standard gamble utility assessments. The probability wheel's primary drawbacks are (1) poor display of small risks of deaths (e.g., less than ten percent) [16] or small increments in risk and (2) a confusing representation of risk to patients unfamiliar with the concept of probability. The design problem for simplifying standard gamble assessment hinges on (1) concretely illustrating small risks of death, (2) reducing the emphasis on abstract representation of probabilities, and (3) avoiding emotional distress on the part of the patient. A prototype design solution is displayed in Figure 4. In the animate graphic, one "person" out of a population of 100 disappears at random for each increase of one percent in risk of death. Reducing the risk of death causes a face figure reappear. Tests of this design are in progress.

Acknowledgements:

The authors were supported by grants HS-00028-01A1 and HS-000028 from the Agency for Health Care Policy Research.

References

1. Levine S. Groog SH. What constitutes quality of life? In: Assessments of quality of life in clinical trials of cardiovascular disease. Le Jacq Publ. 1984, pp. 46-58.
2. von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton: Princeton University Press, 1976.
3. Kaplan RM, Robbit RA, Kressel S, Pollard WE, Gibson BS, Morris JR. Health status: types of validity and the Index of Well-Being. Health Services Research 1976 Winter:478-507.
4. Campbell A, Coverse PE. The quality of American life. New York: Russel Sage Foundations, 1976.
5.Evans RW, Rader B, Manninen DL. The quality of life of hemodialysis recipients treated with recombinant human erythropoietin. JAMA 1990. 263:825-831.
6. Froberg DG, Kane RL. Methodology for measuring health-state preferences--IV progress and a research agenda. J Clin Epidem 198942:675-685.
7.Sox HC, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. Boston: Butterworths, 1988
8. Howard RA. Life and death decisions. In: Howard RA, Matheson JE, ed. The Principles and Practice of Decision Analysis. Menlo Park: Strategic Decisions Group, 1984, pp. 483-506.
9. Howard RA. On fates comparable to death. In: Principles and Practice of Decision Analysis. Menlo Park: Strategic Decisions Group, 1984, pp. 547-573.
10. Torrance GW. Toward a utility theory foundation for health status index models. Health Policy Research 1976 Winter:349-3
11. Torrence G. Utility approach to measuring health-related quality of life. J Chron Dis 1987 40:593-600.
12. Bursztajn H, Hamm RM. The clinical utility of utility assessment. Med Decis Making 198 22(2):161-165.
13. Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes: comparison of assessment methods. Med Decis Making 1984 44:315-329.
14. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981 211:453-458.
15. Hornberger JC, Redelmeier D, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. Journal of Clinical Epidemiology, 1992 45(5):505-12
16. Spetzer CS, Stael von Holstein CS. Probability encoding in decision analysis. In: Howard RA, Matheson JE, ed. The Principles and Practice of Decision Analysis. Menlo Park: Strategic Decisions Group, 1984, pp. 603-625.
17. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science 1974 185:1124-1131.
18. Miller R.G. Simulataneous Statistical Inference. New York: Springer-Verlag, pp. 195-201.
19. Churchill DN, Torrance GW, Taylor DW, Barnes CC, Ludwin D, et. al. Measurement of quality of life in end-stage renal disease: the time trade-off approach. Clinical Invest Med (Canada) 1987 10(1):14-20.